**What Is Claimed Is:**

1. A method for characterizing a document with respect to clusters of conceptually related words, comprising:

receiving the document, wherein the document contains a set of words;

selecting candidate clusters of conceptually related words that are related to the set of words;

wherein the candidate clusters are selected using a model that explains how sets of words are generated from clusters of conceptually related words; and

constructing a set of components to characterize the document, wherein the set of components includes components for candidate clusters, wherein each component indicates a degree to which a corresponding candidate cluster is related to the set of words.

2. The method of claim 1, wherein the model is a probabilistic model, which contains nodes representing random variables for words and for clusters of conceptually related words.

3. The method of claim 2, wherein each component in the set of components indicates a degree to which a corresponding candidate cluster is active in generating the set of words.

4. The method of claim 3,

wherein nodes in the probabilistic model are coupled together by weighted links; and

wherein if a cluster node in the probabilistic model fires, a weighted link from the cluster node to another node can cause the other node to fire.

5. The method of claim 4, wherein if a node has multiple parent nodes that are active, the probability that the node does not fire is the product of the probabilities that links from the active parent nodes do not fire.

6. The method of claim 2, wherein the probabilistic model includes a universal node that is always active and that has weighted links to all cluster nodes.

7. The method of claim 4, wherein selecting the candidate clusters involves:

constructing an evidence tree by starting with terminal nodes associated with the set of words in the document, and following links in the reverse direction to parent cluster nodes;

using the evidence tree to estimate a likelihood that each parent cluster node was active in generating the set of words; and

selecting a parent cluster node to be a candidate cluster node based on its estimated likelihood.

8. The method of claim 7, wherein estimating the likelihood that a given parent node is active in generating the set of words may involve considering:

the unconditional probability that the given parent node is active;

conditional probabilities that the given parent node is active assuming parent nodes of the given parent node are active; and

conditional probabilities that the given parent node is active assuming child nodes of the given parent node are active.

9. The method of claim 8, wherein considering the conditional probabilities involves considering weights on links between nodes.

10. The method of claim 7 wherein estimating the likelihood that a given parent node is active in generating the set of words involves marking terminal nodes during the estimation process to ensure that terminal nodes are not factored into the estimation more than once.

11. The method of claim 7, wherein constructing the evidence tree involves pruning unlikely nodes from the evidence tree.

12. The method of claim 3, wherein during construction of the set of components, the degree to which a candidate cluster is active in generating the set of words is determined by calculating a probability that a candidate cluster is active in generating the set of words.

13. The method of claim 3, wherein during construction of the set of components, the degree to which a candidate cluster is active in generating the set of words is determined by multiplying a probability that a candidate cluster is active in generating the set of words by an activation for the candidate cluster, wherein the activation indicates how many links from the candidate cluster to other nodes are likely to fire.

14. The method of claim 1, wherein constructing the set of components involves normalizing the set of components.

15. The method of claim 3, wherein constructing the set of components involves approximating a probability that a given candidate cluster is active over states of the probabilistic model that could have generated the set of words.

16. The method of claim 15, wherein approximating the probability involves:

selecting states for the probabilistic model that are likely to have generated the set of words in the document; and

considering only selected states while calculating the probability that the given candidate cluster is active.

17. The method of claim 16, wherein selecting a state that is likely to have generated the set of words involves:

randomly selecting a starting state for the probabilistic model; and

performing hill-climbing operations beginning at the starting state to reach a state that is likely to have generated the set of words.

18. The method of claim 17, wherein performing the hill-climbing operations involves periodically changing states of individual candidate clusters without regards to an objective function for the hill-climbing operations to explore states of the probabilistic model that are otherwise unreachable through hill-climbing operations.

19. The method of claim 18, wherein changing a state of an individual candidate cluster involves temporarily fixing the changed state to produce a local optimum for the objective function, which includes the changed state.

20. The method of claim 1, wherein the document can include:

a web page; or

a set of terms from a query.

21. A computer-readable storage medium storing instructions that when executed by a computer cause the computer to perform a method for characterizing a document with respect to clusters of conceptually related words, the method comprising:

receiving the document, wherein the document contains a set of words;

selecting candidate clusters of conceptually related words that are related to the set of words;

wherein the candidate clusters are selected using a model that explains how sets of words are generated from clusters of conceptually related words; and

constructing a set of components to characterize the document, wherein the set of components includes components for candidate clusters, wherein each component indicates a degree to which a corresponding candidate cluster is related to the set of words.

22. The computer-readable storage medium of claim 21, wherein the model is a probabilistic model, which contains nodes representing random variables for words and for clusters of conceptually related words.

23. The computer-readable storage medium of claim 22, wherein each component in the set of components indicates a degree to which a corresponding candidate cluster is active in generating the set of words.

24. The computer-readable storage medium of claim 23,

wherein nodes in the probabilistic model are coupled together by weighted links; and

wherein if a cluster node in the probabilistic model fires, a weighted link from the cluster node to another node can cause the other node to fire.

25. The computer-readable storage medium of claim 24, wherein if a node has multiple parent nodes that are active, the probability that the node does not fire is the product of the probabilities that links from the active parent nodes do not fire.

26. The computer-readable storage medium of claim 22, wherein the probabilistic model includes a universal node that is always active and that has weighted links to all cluster nodes.

27. The computer-readable storage medium of claim 24, wherein selecting the candidate clusters involves:

constructing an evidence tree by starting with terminal nodes associated with the set of words in the document, and following links in the reverse direction to parent cluster nodes;

using the evidence tree to estimate a likelihood that each parent cluster node was active in generating the set of words; and

selecting a parent cluster node to be a candidate cluster node based on its estimated likelihood.

28. The computer-readable storage medium of claim 27, wherein estimating the likelihood that a given parent node is active in generating the set of words may involve considering:

the unconditional probability that the given parent node is active;

conditional probabilities that the given parent node is active assuming parent nodes of the given parent node are active; and

conditional probabilities that the given parent node is active assuming child nodes of the given parent node are active.

29. The computer-readable storage medium of claim 28, wherein considering the conditional probabilities involves considering weights on links between nodes.

30. The computer-readable storage medium of claim 27, wherein estimating the likelihood that a given parent node is active involves marking terminal nodes during the estimation process to ensure that terminal nodes are not factored into the estimation more than once.

31. The computer-readable storage medium of claim 27, wherein constructing the evidence tree involves pruning unlikely nodes from the evidence tree.

32. The computer-readable storage medium of claim 23, wherein during construction of the set of components, the degree to which a candidate cluster is active in generating the set of words is determined by calculating a probability that a candidate cluster is active in generating the set of words.

33. The computer-readable storage medium of claim 23, wherein during construction of the set of components, the degree to which a candidate cluster is active in generating the set of words is determined by multiplying a

probability that a candidate cluster is active in generating the set of words by an activation for the candidate cluster, wherein the activation indicates how many links from the candidate cluster to other nodes are likely to fire.

34. The computer-readable storage medium of claim 21, wherein constructing the set of components involves normalizing the set of components.

35. The computer-readable storage medium of claim 23, wherein constructing the set of components involves approximating a probability that a given candidate cluster is active over states of the probabilistic model that could have generated the set of words.

36. The computer-readable storage medium of claim 35, wherein approximating the probability involves:

selecting states for the probabilistic model that are likely to have generated the set of words in the document; and

considering only selected states while calculating the probability that the given candidate cluster is active.

37. The computer-readable storage medium of claim 36, wherein selecting a state that is likely to have generated the set of words involves:

randomly selecting a starting state for the probabilistic model; and

performing hill-climbing operations beginning at the starting state to reach a state that is likely to have generated the set of words.

38. The computer-readable storage medium of claim 37, wherein performing the hill-climbing operations involves periodically changing states of

individual candidate clusters without regards to an objective function for the hill-climbing operations to explore states of the probabilistic model that are otherwise unreachable through hill-climbing operations.

39. The computer-readable storage medium of claim 38, wherein changing a state of an individual candidate cluster involves temporarily fixing the changed state to produce a local optimum for the objective function, which includes the changed state.

40. The computer-readable storage medium of claim 21, wherein the document can include:

a web page; or

a set of terms from a query.

41. An apparatus for characterizing a document with respect to clusters of conceptually related words, comprising:

a receiving mechanism, configured to receive the document, wherein the document contains a set of words;

a selection mechanism configured to select candidate clusters of conceptually related words that are related to the set of words;

wherein the candidate clusters are selected using a model that explains how sets of words are generated from clusters of conceptually related words; and

a component construction mechanism configured to construct a set of components to characterize the document, wherein the set of components includes components for candidate clusters, wherein each component indicates a degree to which a corresponding candidate cluster is related to the set of words.

42. The apparatus of claim 41, wherein the model is a probabilistic model, which contains nodes representing random variables for words and for clusters of conceptually related words.

43. The apparatus of claim 42, wherein each component in the set of components indicates a degree to which a corresponding candidate cluster is active in generating the set of words.

44. The apparatus of claim 43,

wherein nodes in the probabilistic model are coupled together by weighted links; and

wherein if a cluster node in the probabilistic model fires, a weighted link from the cluster node to another node can cause the other node to fire.

45. The apparatus of claim 44, wherein if a node has multiple parent nodes that are active, the probability that the node does not fire is the product of the probabilities that links from the active parent nodes do not fire.

46. The apparatus of claim 43, wherein the probabilistic model includes a universal node that is always active and that has weighted links to all cluster nodes.

47. The apparatus of claim 44, wherein the selection mechanism is configured to:

construct an evidence tree by starting with terminal nodes associated with the set of words in the document, and following links in the reverse direction to parent cluster nodes;

use the evidence tree to estimate a likelihood that each parent cluster node was active in generating the set of words; and to

select a parent cluster node to be a candidate cluster node based on its estimated likelihood.

48. The apparatus of claim 47, wherein while estimating the likelihood that a given parent node is active in generating the set of words, the selection mechanism is configured to consider at least one of the following:

the unconditional probability that the given parent node is active;

conditional probabilities that the given parent node is active assuming parent nodes of the given parent node are active; and

conditional probabilities that the given parent node is active assuming child nodes of the given parent node are active.

49. The apparatus of claim 48, wherein while considering the conditional probabilities, the selection mechanism is configured to consider weights on links between nodes.

50. The apparatus of claim 47, wherein while estimating the likelihood that a given parent node is active in generating the set of words, the selection mechanism is configure to mark terminal nodes during the estimation process to ensure that terminal nodes are not factored into the estimation more than once.

51. The apparatus of claim 47, wherein while constructing the evidence tree, the selection mechanism is configured to prune unlikely nodes from the evidence tree.

52. The apparatus of claim 43, wherein while constructing a given component in the set of components, the component construction mechanism is configured to determine the degree to which a candidate cluster is active in generating the set of words by calculating a probability that a candidate cluster is active in generating the set of words.

53. The apparatus of claim 43, wherein while constructing a given component in the set of components, the component construction mechanism is configured to determine the degree to which a candidate cluster is active in generating the set of words by multiplying a probability that a candidate cluster is active in generating the set of words by an activation for the candidate cluster, wherein the activation indicates how many links from the candidate cluster to other nodes are likely to fire.

54. The apparatus of claim 41, wherein the component construction mechanism is configured to normalize the set of components.

55. The apparatus of claim 43, wherein the component construction mechanism is configured to approximate a probability that a given candidate cluster is active over states of the probabilistic model that could have generated the set of words.

56. The apparatus of claim 55, wherein while approximating the probability, the component construction mechanism is configured to:

select states for the probabilistic model that are likely to have generated the set of words in the document; and to

consider only selected states while calculating the probability that the given candidate cluster is active.

57. The apparatus of claim 56, wherein while selecting a state that is likely to have generated the set of words, the component construction mechanism is configured to:

randomly select a starting state for the probabilistic model; and to

perform hill-climbing operations beginning at the starting state to reach a state that is likely to have generated the set of words.

58. The apparatus of claim 58, wherein while performing the hill-climbing operations, the component construction mechanism is configured to periodically change states of individual candidate clusters without regards to an objective function for the hill-climbing operations to explore states of the probabilistic model that are otherwise unreachable through hill-climbing operations.

59. The apparatus of claim 58, wherein while changing a state of an individual candidate cluster, the component construction mechanism is configured to temporarily fix the changed state to produce a local optimum for the objective function, which includes the changed state.

60. The apparatus of claim 41, wherein the document can include:

a web page; or

a set of terms from a query.

61. A computer-readable storage medium containing a data structure that facilitates characterizing a document with respect to clusters of conceptually related words, the data structure comprising:

a probabilistic model that contains nodes representing random variables for words and for clusters of conceptually related words;

wherein nodes in the probabilistic model are coupled together by weighted links;

wherein if a cluster node in the probabilistic model fires, a weighted link from the cluster node to another node can cause the other node to fire; and

wherein the other code can be associated with a word or a cluster.

62. The computer-readable storage medium of claim 61, wherein the probabilistic model includes a universal node that is always active and that has weighted links to all cluster nodes.